Applications of Finite deFinetti
Style Theorems to Linear Models

by

Persi Diaconis[1], Morris Eaton[2], and Steffen Lauritzen
Harvard University, University of Minnesota
and Aalborg University
Technical Report No. 502
October 1987

The authors would like to thank the Danish Research Council for financial support for portions of this research work.

---

[1] This work was supported in part by National Science Foundation Grant No. DMS-86-00235.

[2] This work was suported in part by National Science Foundation Grant No. DMS-83-19924

Summary

In this paper, the implications of invariance assumptions on statistical models is studied via finite forms of deFinetti-type Theorems. Bounds on the variation distance between a given invariant model and mixtures of certain standard statistical models are provided for matrix models which arise in multivariate analysis and for univariate linear models. These bounds are illustrated in the case of simple linear regression.

§1 Introduction:

The results in this paper are concerned with the implications of invariance and extendability assumptions for statistical models. Here is an example from Diaconis and Freedman (1980). Suppose a random vector $X = (X_1, \dots, X_k)$ has coordinates which are 0 or 1 and suppose the distribution of X is exchangeable - that is, $(X_1, \dots, X_k)$ and $(X_{\pi(1)}, \dots, X_{\pi(k)})$ have the same distribution for all permutations π. The distribution of X is n-extendable if there exists a vector $Y = (Y_1, \dots, Y_n)$ such that Y is exchangeable and $(X_1, \dots, X_k)$ has the same distribution as $(Y_1, \dots, Y_k)$.

The classical deFinetti Theorem asserts that if P, the probability measure of X, is n-extendable for all n = k+1, k+2..., then P has the representation

$$P = \int_0^1 Q_\alpha \mu(d\alpha)$$

where Q_α is the probability measure of $Z_1, \dots, Z_k$ which are iid Bernoulli random variables with success probability α, and μ is some probability measure on [0,1]. In other words, P is n-extendable for all n iff P is a mixture of Q_α's.

When X is n-extendable for a fixed n, then deFinetti's Theorem fails. However Diaconis and Freedman (1980) established the following "finite version" of deFinetti's Theorem. Let $||\cdot||$ denote variation distance for measures (defined precisely in the next section). Assume X is n-extendable, and let P denote the distribution of X. Then, Diaconis and Freedman (1980) showed that there exists a measure μ on [0,1] such that

$$P_\mu = \int_0^1 Q_\alpha \mu(d\alpha)$$

satisfies

$$||P-P_\mu|| \le \frac{4\ k}{n} .$$

In other words, P is within 4k/n of some mixture of iid Bernoulli's where closeness is measured by variation distance.

The above result shows that for 0-1 random variables, exchangeability (invariance under permutations) together with extendability imply that the statistical model for the original variables is close to some mixture of iid Bernoullis. Thus one has an approximation to the distribution of the original varaiables. A second example of this type, involving the orthogonal group (as opposed to the permutation group) and scale mixtures of iid mean zero normal variables is given in the next section. This example comes from Diaconis and Freedman (1986) where a host of other examples of this type are also given. However most of the examples in Diaconis and Freedman (1986) are phrased in terms of sufficient statistics rather than invariance.

This paper contains two types of examples for which finite deFinetti-style theorems are proved. The first type concerns models involving the multivariate normal distribution. Example 2.1 establishes a finite version of a multivariate result proved in Dawid (1978). Example 2.2 gives both a generalization and a finite version of a result in Smith (1981). A general method of proof on which both of these examples rely is also outlined in Section 2.

The second type of example discussed in Section 3, involves univariate linear models. Here, a general finite theorem is established and applied to the linear regression model. The method of proof applies to multivariate as well as univariate linear models, but we have not completed the detailed calculations for these examples. Work is currently under way on further linear model examples--both fixed and random effects models.

It should be mentioned that associated with every finite theorem we know, there is a corresponding "infinite" theorem which results from assuming extendability for "arbitrarily large n". In many cases, the "infinite theorem" was proved first and suggested some finite theorem. There is a very rich literature on the infinite theorems and their relationship to sufficiency, statistical modelling and the notion of repetitive structures--see Lauritzen (1982) and Diaconis and Freedman (1984) for a discussion and further references.

The appendix to this paper contains a number of technical results concerning random orthogonal matrices. A key inequality provides an upper bound on the variation distance between the following two distributions on $r \times s$ real matrices:

(i) Let U be an $n \times n$ random orthogonal matrix (uniform distribution on the group O_n of $n \times n$ orthogonal matrices). Let $\Delta : r \times s$ be the upper left hand corner of U, and let P_1 be the distribution of Δ.

(ii) Let $V : r \times s$ be an $r \times s$ matrix of iid $N(0, n^{-1})$ random variables and let P_2

be the distribution of V.

Fix $\gamma \in (0,1)$. For $(r + s + 2)/n \le \gamma$, Proposition A.3 shows that the variation distance between P_1 and P_2 is bounded above by

$$2a(\gamma)\,\frac{r + s + 2}{n}$$

where the constant $a(\gamma)$ is given in the Appendix. This result is used in all of the examples in this paper.

In short, here is an outline of the remainder of this paper. Section 2 contains a general method of obtaining finite theorems under invariance and extendability assumptions. This method is used to establish two finite theorems for some multivariate analysis models of interest. In section 3, a general finite theorem is proved for univariate linear models and is applied to the simple linear regression model. The appendix contains technical results regarding random orthogonal matrices.

§2 Some Multivariate Examples.

In this section the implications of invariance and extendability assumptions are investigated for some univariate and multivariate examples. Before describing the general method of proof, it is useful to consider an example. Suppose Y is a random vector in R^n which has an orthogonally invariant distribution. More precisely, let 0_n be the group of n×n orthogonal matrices and assume

$$\mathcal{L}(Y) = \mathcal{L}(gY) \quad , \quad g \in 0_n \tag{2.1}$$

where $\mathcal{L}(\cdot)$ denotes the law of "•". Consider the k-dimensional random vector $X \in R^k$ which is the first k coordinates of Y. The problem is to say something about $\mathcal{L}(X)$. Here is a result of Diaconis and Freedman (1986). Let $N_k(0,\sigma^2 I_k)$ denote the normal distribution on R^k with mean 0 and covariance $\sigma^2 I_k$. Given any probability measure α on $[0,\infty)$, let S_α denote be the mixture on σ, with respect to α, of $N_k(0,\sigma^2 I_k)$. Then there exists an S_α such that the variation distance between $\mathcal{L}(X)$ and S_α is bounded above by a constant times k/n when k is at most a fraction of n.

One proof of this result runs as follows. Since $\mathcal{L}(Y)$ is orthogonally invariant, Y can be represented (in distribution) as RV where R is a non-negative random variable independent of V and V has a uniform distribution on the sphere of radius one in R^n. Thus, X has the same distribution as RW where $W \in R^k$ is the first k coordinates of V. It is easy to show that $\underset{\sim}{E}W = 0$ and measurable space $(\underset{\sim}{X}, \mathcal{B}_1)$ and a measurable map π: $\underset{\sim}{X}_2 \dashrightarrow \underset{\sim}{X}_1$, let

$$\mathcal{P}_{12} = \{\pi P | P \in \mathcal{P}_2\} , \tag{2.2}$$

where πP is defined by

$$(\pi P)(B_1) = P(\pi^{-1}(B_1)) \quad , \quad B_1 \in \mathcal{B}_1 .$$

In most cases, it is useful to think of π as a "projection" so $\mathcal{P}_{12}$ is the set of all the projected invariant measures.

Let ν denote the unique invariant probability measure on G. A random element of G, say U, is <u>uniform on G</u> if $\mathcal{L}(U) = \nu$. Given $y \in \underset{\sim}{X}_2$ and U uniform on G, Uy is the point y acted on by U, so Uy is a random element in $\underset{\sim}{X}_2$. Let $\nu_y = \mathcal{L}(Uy)$, and for each y consider an "approximation" Q_y to ν_y. Thus for each y, Q_y is a probability measure on $(\underset{\sim}{X}_2, \mathcal{B}_2)$. The set of all mixtures of $\{Q_y | y \in \underset{\sim}{X}_2\}$ is denoted by $\underset{\sim}{S}$. Proposition 2.1 below shows that for each $\pi P \in \mathcal{P}_{12}$, there is an element in $\pi\underset{\sim}{S}$ whose variation distance from πP is at most

$$d = \sup_y ||\pi\nu_y - \pi Q_y|| \tag{2.3}$$

where $||\cdot||$ is variation distance.

<u>Remark 2.1</u>: To see how the first example relates to the abstract formulation, take $\underset{\sim}{X}_2 = R^n$, $\underset{\sim}{X}_1 = R^k$ and $G = O_n$ so $\mathcal{P}_2$ is all orthogonally invariant

$\mathrm{Cov}(W) = n^{-1}I_k$. Thus, with R fixed, $X = RW$ has mean 0 and covariance $n^{-1}R^2I_k$. Let Z be $N(0,I_k)$ and approximate the distribution of X by the distribution of $n^{-1/2}RZ$ with R and Z independent. Because variation distance, say $||\cdot||$, is convex, it follows that

$$||\mathcal{L}(RW) - \mathcal{L}(n^{-1/2}RZ)|| \leq$$

$$\sup_t ||\mathcal{L}(tW) - \mathcal{L}(n^{-1/2}tZ)|| =$$

$$||\mathcal{L}(W) - \mathcal{L}(n^{-1/2}Z)||$$

The final equality results from the observation that the scale factor t does not affect variation distance. The last step in the proof, which consists of some analysis, bounds

$$||\mathcal{L}(W) - \mathcal{L}(n^{-1/2}Z||$$

above by using the explicit expressions for $\mathcal{L}(W)$ and the normal distribution. This yields the claimed bound. It is interesting that the S_α appears explicitly in the proof - namely $S_\alpha = \mathcal{L}(n^{-1/2}RZ)$ so α is the distribution of $n^{-1/2}R$.

Before turning to the examples in this section, it is useful to abstract certain portions of the argument in the example above. Consider a measurable space $(\underset{\sim}{X}_2, \mathcal{B}_2)$ which is acted on measurably by a compact topological group G. Let $\mathcal{P}_2$ be all the G-invariant probability measures on $(\underset{\sim}{X}_2, \mathcal{B}_2)$. Given a second probabilities on R^n. The projection π is given by the matrix

$$\pi = (I_k 0) : k \times n$$

so $\pi Y = X$ in the notation of the example. A unif[illegible]m $n \times n$ orthogonal matrix, and $Uy \in R^n$ is a random vector with a uniform distribution on the sphere of radius $||y||$. With $\nu_y = \mathcal{L}(Uy)$, the "approximation" Q_y was chosen to be normal with the same mean and covariance as Uy. The proposition below is the abstract version of the inequality in (2.1). □

<u>Proposition 2.1</u>: For each $P \in \mathcal{P}_2$,

$$(2.4) \qquad \delta = \inf_{Q \in \underset{\sim}{S}} ||\pi P - \pi Q|| \leq d$$

where d is given in (2.3).

<u>Proof</u>: Given any two probabilities ζ_1 and ζ_2 on $\underset{\sim}{X}_1$, the variation distance between ζ_1 and ζ_2 is

$$(2.5) \qquad ||\zeta_1 - \zeta_2|| = \sup_{|f| \leq 1} \left| \int f(x)\zeta_1(dx) - \int f(x)\zeta_2(dx) \right|$$

where the sup ranges over all measurable functions bounded in absolute value by 1. For such a function and $g \in G$, the definition of πP and the invariance of P yields

$$(2.6) \qquad \int_{\underset{\sim}{X}_1} f(x)(\pi P)(dx) = \int_{\underset{\sim}{X}_2} f(\pi y)P(dy) = \int_{\underset{\sim}{X}_2} f(\pi gy)P(dy).$$

Averaging (2.6) with respect to ν gives

$$(2.7) \qquad \int_{\underset{\sim}{X}_1} f(x)(\pi P)(dx) = \int_{\underset{\sim}{X}_2} \int_G f(\pi gy)\ \nu(dg)\ P(dy)\ .$$

Because Q is an average (over y) of the Q_y's, Q can be written

$$Q = \int_{\underset{\sim}{X}_2} Q_y H(dy)$$

where H is a probability on $\underset{\sim}{X}_2$. Thus

$$(2.8) \qquad \int_{\underset{\sim}{X}_1} f(x)\ (\pi Q)(dx) = \int_{\underset{\sim}{X}_2} f(\pi z)Q(dz) = \int_{\underset{\sim}{X}_2} \int_{\underset{\sim}{X}_2} f(\pi z)\ Q_y(dz)\ H(dy)\ .$$

Using (2.5), (2.7) and (2.8) for any $Q \in S$ gives

$$(2.9) \qquad ||\pi P - \pi Q|| =$$
$$\sup_{|f|\le 1} \left| \int_{\underset{\sim}{X}_2} \int_G f(\pi gy)\ \nu(dg)P(dy) - \int_{\underset{\sim}{X}_2}\int_{\underset{\sim}{X}_2} f(\pi z)Q_y(dx)H(dy) \right|$$

Taking H = P in (2.9) shows that δ in (2.4) satisfies

$$(2.10) \qquad \delta \le \sup_{|f|\le 1}\left|\int_{\underset{\sim}{X}_2}\int_G f(\pi g y)\ \nu(dg)P(dy)-\int_{\underset{\sim}{X}_2}\int_{\underset{\sim}{X}_2} f(\pi z)Q_y(dz)P(dy)\right| =$$

$$\sup_{|f|\le 1}\left|\int_{\underset{\sim}{X}_2}\left[\int_G f(\pi g y)\ \nu(dg) - \int_{\underset{\sim}{X}_2} f(\pi z)Q_y(dz)\right] P(dy)\right| \le$$

$$\sup_{|f|\le 1}\ \sup_{y}\ \left|\int_G f(\pi g y)\nu(dg) - \int_{\underset{\sim}{X}_2} f(\pi z)Q_y(dz)\right| =$$

$$\sup_{y}\ \sup_{|f|\le 1}\ \left|\int_G f(\pi g y)\nu(dg) - \int_{\underset{\sim}{X}_2} f(\pi z)Q_y(dz)\right| \ .$$

But, by definition of ν_y,

$$(2.11) \qquad \int_G f(\pi g y)\nu(dg) = \int_{\underset{\sim}{X}_1} f(x)(\pi\nu_y)(dx) \quad .$$

Also, by definition,

$$(2.12) \qquad \int_{\underset{\sim}{X}_2} f(\pi z)Q_y(dz) = \int_{\underset{\sim}{X}_1} f(x)(\pi Q_y)(dx) \quad .$$

Substituting (2.11) and (2.12) into the last expression in (2.10) yields

$$\delta \le \sup_{y}\ \sup_{|f|\le 1}\ \left|\int_{\underset{\sim}{X}_1} f(x)(\pi\nu_y)dx - \int_{\underset{\sim}{X}_1} f(x)(\pi Q_y)(dx)\right|$$

which, via (2.5), gives

(2.13) $$\delta \le \sup_y ||\pi\nu_y - \pi Q_y|| = d$$

and the desired result. □

Here is the multivariate version of the example discussed at the beginning of this section.

Example 2.1: This example is related to a result in Dawid (1978) and arises in the following way. Suppose $X:k\times p$ is a random matrix with a left orthogonally invariant distribution --that is,

(2.14) $$\mathcal{L}(X) = \mathcal{L}(\Gamma X) \ , \ \Gamma \in O_k \ .$$

For example, X might be $N(0, I_k \otimes \Sigma)$ or perhaps a mixture over Σ of a $N(0, I_k \otimes \Sigma)$ distribution. Here, Σ is a $p\times p$ non-negative definite matrix. $\mathcal{L}(X)$ is extendable if for each $n > k$, there is a random matrix $Y^{(n)}: n \times p$ such that

(2.15) $$\begin{cases} \text{(i) } \mathcal{L}(Y^{(n)}) = \mathcal{L}(\Gamma Y^{(n)}) \ , \quad \Gamma \in O_n \\ \text{(ii) } \mathcal{L}(\pi^{(n)} Y^{(n)}) = \mathcal{L}(X) \end{cases}$$

where $\pi^{(n)}$ is the matrix

(2.16) $$\pi^{(n)} = (I_k 0): k\times n.$$

Proposition 2.2 (Dawid (1978)). Suppose $\mathcal{L}(X)$ satisfies (3.14). Then $\mathcal{L}(X)$ is extendable iff $\mathcal{L}(X)$ is a mixture over Σ of the $N(0, I_k \otimes \Sigma)$ distribution.

The question addressed in this example is "what happens if for some fixed n, there exists a $Y^{(n)}$ satisfying (i) and (ii)?" To formulate the answer, let $\underset{\sim}{S}$ be all distributions which are mixtures over Σ of a $N(0, I_k \otimes \Sigma)$ distribution.

Proposition 2.3: Fix $n > k$. Suppose $Y^{(n)}: n \times p$ satisfies (i) of (2.15) and let $X = \pi^{(n)} Y^{(n)}$ where $\pi^{(n)}$ is given in (2.16). Then there is an $S \in \underset{\sim}{S}$ such that the variation distance between $\mathcal{L}(X)$ and S is bounded by a constant, $2a(\gamma)$, times $(k + p + 2)/n$ as long as $(k + p + 2)/n \leq \gamma$ (here γ is fixed, $0 < \gamma < 1$). The constant $a(\gamma)$ is given in Proposition A.3.

Proof: The idea of the proof is to first apply Proposition 3.1 and then bound the number d in (2.3) by using the results in the Appendix. In Proposition 2.1 take $\underset{\sim}{X}_2$ to be the set of all $n \times p$ real matrices and take π to $\pi^{(n)}$ as defined in (2.16). Thus $\underset{\sim}{X}_1$ is the set of $k \times p$ real matrices. The group O_n is G so U is uniform on $\underset{\sim}{G}$ means U is a "random" $n \times n$ orthogonal matrix. Now for $y \in \underset{\sim}{X}_2$, $\pi^{(n)} \nu_y$ is the probability law of

$$V = \pi^{(n)} U y \quad , \tag{2.17}$$

which has mean 0 and covariance

$$n^{-1} I_k \otimes y' y \quad . \tag{2.18}$$

In this example Q_y is the $N(0,n^{-1}I_n \otimes y'y)$ distribution so

$$\pi^{(n)}Q_y = N(0,\ n^{-1}\ I_k \otimes y'y) \quad . \tag{2.19}$$

Thus $\pi^{(n)}Q_y$ has been completely specified. To describe $\pi^{(n)}\nu_y$ more fully, first write $y{:}n\times p$ as

$$y = \Gamma\begin{pmatrix} I_p \\ 0 \end{pmatrix}(y'y)^{1/2} \tag{2.20}$$

where $\Gamma \in O_n$ and $(y'y)^{1/2}$ denotes the symmetric square root of $y'y{:}p\times p$. (That such a representation exists is well known--for example, see Example 1.11 in Eaton (1983)). Substituting (2.20) in (2.17) gives

$$V = \pi^{(n)}U\Gamma\begin{pmatrix} I_p \\ 0 \end{pmatrix}(y'y)^{1/2} \quad . \tag{2.21}$$

Since U is uniform on O_n, $\mathcal{L}(U) = \mathcal{L}(U\Gamma)$. Setting

$$\Delta = \pi^{(n)}U\Gamma\begin{pmatrix} I_p \\ 0 \end{pmatrix}\ :\ k\times p \tag{2.22}$$

it follows that Δ is distributed as the upper left $k\times p$ corner of a random $n\times n$ orthogonal matrix. Thus

$$\mathcal{L}(V) = \mathcal{L}(\Delta(y'y)^{1/2}) = \pi^{(n)}\nu_y \ .$$

Further, if Z is $N(0, n^{-1}I_k \times I_p)$, then

$$\mathcal{L}(Z(y'y)^{1/2}) = \pi^{(n)}Q_y \quad .$$

Hence

$$(2.23) \qquad d = \sup_y || \pi^{(n)}\nu_y - \pi^{(n)}Q_y|| =$$

$$\sup_y ||\mathcal{L}(\Delta(y'y)^{1/2}) - \mathcal{L}(Z(y'y)^{1/2})|| =$$

$$||\mathcal{L}(\Delta) - \mathcal{L}(Z)|| \ .$$

The last equality in (2.23) is a consequence of inequality (A.35) in the appendix and the fact that for $y'y = I_p$, there is actually equality. Now, the variation distance between $\mathcal{L}(\Delta)$ and $\mathcal{L}(Z)$ is bounded in Proposition A.3 (with $r = k$, $s = p$) and this yields the claimed result. □

It should be noted that the sup over y in (2.3) was calculated explicitly in this example. The same situation occurs in most of the examples. This completes Example 2.1. □

Example 2.2: The motivation for this example comes from Smith (1981) whose result can be described as follows. For each integer m, let $O_m(e)$ be the group of m×m orthogonal matrices Γ such that $\Gamma e = e$ where e is the vector of all ones in R^m (the dependence of e on m is suppressed). Consider a random vector X in R^k whose distribution is $O_k(e)$ invariant--that is

(2.24) $$\mathcal{L}(X) = \mathcal{L}(\Gamma X), \quad \Gamma \in O_k(e).$$

For example, if the coordinates of X are independent normals with the same mean and variance--say

$$\mathcal{L}(X) = N_k(\mu e, \sigma^2 I_k),$$

then (2.24) holds. $\mathcal{L}(X)$ is <u>extendable</u> if for each $n > k$, there is a random vector $Y^{(n)} \in R^n$ such that

(i) $\mathcal{L}(Y^{(n)}) = \mathcal{L}(\Gamma Y^{(n)})$, $\Gamma \in O_n(e)$.

(ii) $\mathcal{L}(\pi^{(n)} Y^{(n)}) = \mathcal{L}(X)$

where

(2.25) $$\pi^{(n)} = (I_k 0): k\times n.$$

Let $\underset{\sim}{S}_k$ be all distributions on R^k which are mixtures, over (μ,σ), of the $N_k(\mu e,\sigma^2 I_k)$ distribution.

<u>Proposition 2.4</u> (Smith (1981)): Let X be a random vector in R^k. Then $\mathcal{L}(X)$ is extendable iff $\mathcal{L}(X) \in \underset{\sim}{S}_k$.

In this example, a multivariate version of Proposition 2.4 is given which is similar in spirit and content to Propositon 2.3. As in the previous example, fix n and consider a random matrix $Y:n\times p$ such that

$$\mathcal{L}(Y) = \mathcal{L}(\Gamma Y) \quad , \quad \Gamma \in 0_n(e) \ . \tag{2.26}$$

With $\pi^{(n)}$ given by (2.25) set

$$X = \pi^{(n)}Y\colon\ k\times p. \tag{2.27}$$

Given a vector $\mu \in R^p$ and a $p\times p$ non-negative definite matrix Σ, let

$$N(e\mu',\ I_k \otimes \Sigma) \tag{2.28}$$

denote the normal distribution on $k\times p$ real matrices with mean $e\mu'(e \in R^k)$ and covariance $I_k\otimes\Sigma$. Also, let $\underset{\sim}{S}$ denote the mixture, over $(\mu,\ \Sigma)$, of these normal distributions. Essentially, the main result of this example shows that there is an $S \in \underset{\sim}{S}$ such that the variation distance between $\mathcal{L}(X)$ and S is bounded above by a constant times $(k + p + 2)/n$.

Proposition 2.5: Let $Y:n\times p$ satisfy (2.26) and define X by (2.27). Then there is a (μ, Σ) mixture of the distributions in (3.28) whose variation distance from $\mathcal{L}(X)$ is bounded above by $2a(\gamma)[(p + k + 2)(n-1)^{-1} + kn^{-1}]$ as long as $(p + k + 2)(n-1)^{-1} \le \gamma < 1$ where $a(\gamma)$ is given in Proposition A.3.

Proof: The proof is a variation of the proof of Proposition 3.3. First, the distribution $\pi^{(n)}\nu_y$ needs to be described. Let U: $n\times n$ be uniform on $O_n(e)$ so that

$$\pi^{(n)}\nu_y = \mathcal{L}(\pi^{(n)}Uy)$$

where y is a fixed $n\times p$ matrix. The orthogonal projection onto the one dimensional subspace of R^n spanned by the vector of ones is

$$P = \frac{1}{n}ee'$$

where $e \in R^n$. Thus $Q = I-P$ is the orthogonal projection onto the orthogonal complement of this subspace. Proposition A.1 implies that

$$\mathcal{L}(\pi^{(n)}Uy) = \mathcal{L}(A^{1/2}\Delta B^{1/2} + \pi^{(n)}Py) \tag{2.29}$$

where

$$(2.30)\quad \begin{cases} \text{(i)} \quad A = \pi^{(n)}Q\pi^{(n)\prime} : k\times k \\ \text{(ii)} \quad B = y'Qy : p\times p \\ \text{(iii)} \ \Delta\text{: } k\times p \text{ has the distribution of the } k\times p \text{ upper left block of an} \\ \qquad (n-1)\times(n-1) \text{ orthogonal matrix uniform on } O_{n-1}. \end{cases}$$

Now, let $Z:k\times p$ be $N(0, (n-1)^{-1} I_k \otimes I_p)$. Then for fixed $y:n\times p$,

$$(2.31)\qquad \mathcal{L}(ZB^{1/2} + \pi^{(n)}Py) = N(e\mu', I_k \otimes ((n-1)^{-1}B))$$

where e is the vector of ones in R^k and $\mu \in R^p$ is defined so that

$$\pi^{(n)}Py = e\mu'.$$

Choosing Q_y to be the distribution in (2.31), Proposition 2.1 shows that the variations distance, δ, between $\mathcal{L}(X)$ and the closest mixtures of the distributions in (2.28) satisfies

$$\delta \le \sup_y ||\mathcal{L}(A^{1/2}\Delta B^{1/2} + e\mu') - \mathcal{L}(ZB^{1/2} + e\mu')|| \le$$

$$||\mathcal{L}(A^{1/2}\Delta) - \mathcal{L}(Z)||$$

where inequality (A.35) has been used.

Thus,

$$\delta \le || \mathcal{L}(A^{1/2}\Delta) - \mathcal{L}(A^{1/2}Z) + \mathcal{L}(A^{1/2}Z) - \mathcal{L}(Z)|| \le$$
$$||\mathcal{L}(A^{1/2}\Delta) - \mathcal{L}(A^{1/2}Z|| + ||\mathcal{L}(A^{1/2}Z) - \mathcal{L}(Z)|| =$$
$$||\mathcal{L}(\Delta) - \mathcal{L}(Z)|| + ||\mathcal{L}(A^{1/2}Z) - \mathcal{L}(Z)|| \ .$$

Now Proposition (A.3) bounds $||\mathcal{L}(\Delta) - \mathcal{L}(Z)||$ by $2a(\gamma)(p + k + 2)(n-1)^{-1}$ as long as $(p + k + 2)(n-1)^{-1} \le \gamma < 1$ and $p + k \le n-4$. A direct application of Proposition A.4 shows that

$$||\mathcal{L}(A^{1/2}Z) - \mathcal{L}(Z)|| \le 2[(\det A)^{-p/2} - 1].$$

But

$$A = \pi^{(n)}Q\pi^{(n)'} = I_k - \frac{1}{n}ee'$$

so

$$\det A = 1 - \frac{k}{n}$$

Since $k/n \le \gamma$, the argument in the proof of Proposition A.3 shows that

$$(1 - \frac{k}{n})^{-p/2} - 1 \le b(\gamma)\frac{k}{n}$$

where

$$b(\gamma) = \frac{(1 - \gamma)^{-p/2} - 1}{\gamma} \le a(\gamma) \quad .$$

Hence

$$2((\det A)^{-p/2} - 1) \le 2a(\gamma)\ \frac{k}{n}$$

which completes the proof. $\square$

§ 3 Univariate Linear Models:

We begin this section with an example from simple linear regression. The regression problem forms the prototype of most other simple linear model examples that we know.

Example 3.1: Consider a data vector $Y_1 \epsilon R^k$. Assume a model for Y_1 of the form

$$Y_1 = X_1\beta + \epsilon_1$$

where

$$X_1 = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_k \end{pmatrix} : k\times 2$$

is known and has rank 2. The vector of parameters $\beta \epsilon R^2$ is unknown. Typically the error vector ϵ_1 is assumed to be $N_k(0,\sigma^2 I_k)$, but here it is assumed that ϵ_1 has an 0_k-invariant distribution.

Now, imagine that a larger experiment had been done yielding $Y \epsilon R^n$ $(n > k)$ with a model

$$Y = X\beta + \epsilon$$

It is assumed that the original experiment is "imbedded" in this larger experiment. To be explicit, assume that

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} , \; X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} , \; \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$$

where Y_1, X_1 and ϵ_1 come from the first experiment.

The main result in this section describes the implications of assuming that ϵ has an O_n-invariant distributuion. To state this result for the example at hand, let P_1 denote the distribution of Y_1. Further let $P_{\mu,\sigma}$ denote a normal distribution on R^k with a mean vector μ of the form

$$\mu = X_1\beta \quad , \quad \beta \epsilon R^2 \; ,$$

and a covariance

$$\sigma^2 I_k \; , \; \sigma \geq 0 \; .$$

Let M be the two dimensional subspace of possible values of μ. Under the assumption that ϵ has an O_n-invariant distribution, the main result of this section shows that there exists a distribution P^* which is a mixture of the family

$$\{P_{\mu,\sigma} | \mu \epsilon M, \; \sigma \geq 0\}$$

such that the variation distance between P_1 and P^* is bounded above by two times

$$B_{k,n}(X) = c\,\frac{k+3}{n-2} + 2\left[\left(\frac{\det(X'X)}{(\det(X_2'X_2))}\right)^{1/2} - 1\right]$$

as long as k is at most a fraction of n. Here, c is a constant independent of k,n and the design matrix X. The exact value of c is given in Example 3.2 where we discuss the regression example further.

The implication of this result is that if $B_{k,n}(X)$ is small [so n is large compared to k and $\det(X'X) \simeq \det(X_2'X_2)$], then to a close approximation, the model P_1 arose by first drawing (μ,σ) according to some mixing distribution, and then Y_1 was drawn form a $N_k(\mu,\sigma^2 I_k)$ distribution. This ends Example 3.1. □

To begin the general discussion, consider a finite dimensional real inner product space (V, (.,.)) and let M be a fixed linear subspace of V. The orthogonal group of (V, (.,.)), say O_V, is the group of all linear transformations on V to V which preserve the inner product (.,.) - that is, $g \in O_V$ iff g is linear and

$$(gx,\ gy) = (x,y) \text{ for all } x,\ y \in V.$$

The subgroup $O_V(M)$ is defined by

$$O_V(M) = \{g \mid g \in O_V,\ gx = x \text{ for all } x \in M\}\ .$$

Let $\mathcal{P}$ be all the probability distributions on V which are invariant under the compact group $G = O_V(M)$.

Remark 3.1: In the regression example, $(V, (\cdot,\cdot))$ is R^n with the usual inner product, M is the column space of the design matrix X and the distribution of the observation vector $Y \in R^n$ is invariant under $O_V(M)$. The vector Y_1 is obtained by projecting Y to a smaller space. This aspect of the problem is treated next. □

Now, consider a second inner product space $(V_1, (\cdot,\cdot)_1)$ and let π be a linear transformation on V to V_1 which satisfies

$$\pi\pi' = I_1 .$$

Here π' is the adjoint of π and I_1 is the identity transformation on V_1. Set $Y_1 = \pi Y$ and set

$$\mathcal{P}_1 = \{\pi P | P \in \mathcal{P}\} .$$

Thus, if $\mathcal{L}(Y) = P$, then $\mathcal{L}(\pi Y) = \pi P$. The linear subspace $M_1 \subseteq V$ defined by

$$M_1 = \pi(M)$$

is just the image of M under the linear map π.

To describe the main result of this section, let $\underset{\sim}{S}$ be all distributions on V_1 which, for some probability measure ξ, have the form

$$\int_{M_1}\int_0^\infty N(\mu,\sigma^2 I_1)\ \xi(d\mu,\ d\sigma)$$

where $N(\mu,\ \sigma^2 I_1)$ denotes the normal distribution on V_1 with mean vector μ and covariance $\sigma^2 I_1$. Thus $S \in \underset{\sim}{S}$ means that S is a mixture of normal distributions with mean vectors in M_1 and covariances of the form $\sigma^2 I_1$. Given $P \in \mathcal{P}$, Theorem 3.1 gives an upper bound on the variation distance between πP and the closest element in $\underset{\sim}{S}$ to πP. More precisely, a bound B is derived for

$$\inf_{S \in \underset{\sim}{S}} ||\pi P - S|| \quad .$$

The bound B does not depend on P, but does depend on the dimensions of V, M, V_1, and the relationship between M and the projection π. All of this is made precise in the discussion that follows.

Let $n = \dim V$, $m = \dim M$, $k = \dim V_1$, and let Q be the orthogonal projection onto the orthogonal completement of M.

<u>Theorem 3.1</u> Let $\gamma \in (0,1)$ be fixed and assume $k + 3 \le \gamma\ (n-m)$. Assume that $\pi Q \pi'$ is non-singular. Then given $P \in \mathcal{P}$, there exists an $S \in \underset{\sim}{S}$ such that

$$||\pi P - S|| \le 2B(k,\ n-m,\ \pi)$$

where

$$B(k,\ n\text{-}m,\ \pi) = a(\gamma)\frac{k+3}{n-m} + \left[\left(\det\pi Q\pi'\right)^{-1/2} - 1\right] .$$

The constant $a(\gamma)$ is given in Proposition A.3.

<u>Proof</u>: We use the method of proof described in the proof of Proposition 2.1. Let U be uniform on the group $O_V(M)$. We will obtain an upper bound on d in (2.13). For $y \in V$, write

$$y = y_1 + y_2$$

where $y_1 \in M$ and $y_2 \in M^{\perp}$. Since $U \in O_V(M)$,

$$Uy = y_1 + Uy_2$$

and thus

$$\pi Uy = \pi y_1 + \pi Uy_2 \quad .$$

Fix a point $y_0 \in M^{\perp}$ such that $(y_0, y_0) = 1$. The transitivity of $O_V(M)$ on

$$\{y \mid (y,y) = 1,\ y \in M^{\perp}\}$$

implies that πUy_2 has the same distribution as

$$|y_2| \ \pi U y_0 .$$

where $|y_2| = (y_2, y_2)^{1/2}$. The results in Lemma A.1 imply that

$$E\pi Uy = \pi y_1 \in M_1$$

and

$$\mathrm{Cov}(\pi Uy) = (n-m)^{-1} \ |y_2|^2 \pi Q \pi' \ .$$

Thus, we pick our approximation πQ_y to be

$$N(\pi y_1, \ (n-m)^{-1} \ |y_2|^2 \ \pi Q \pi') \ .$$

Thus, to bound the variation distance between

$$\pi\nu_y = \mathcal{L}(\pi U y_2 + \pi y_1)$$

and the closest element $S \in \underset{\sim}{S}$, it is sufficient to bound

(3.1) $$||\pi\nu_y - N(\pi y_1, \ (n-m)^{-1}|y_2|^2 I_1|| \ .$$

Using the triangle inequality, (3.1) is bounded above by

$$(3.2) \qquad ||\pi\nu_y - N(\pi y_1,(n-m)^{-1}|y_2|^2\pi Q\pi')|| +$$

$$||N(\pi y_1,\ (n-m)^{-1}|y_2|^2\pi Q\pi') - N(\pi y_1,\ (n-m)^{-1}|y_2|^2 I_1)||$$

A direct application of Proposition A.1 (with $\alpha = \pi$ and $\beta' = y_0$) coupled with Proposition A.3 (with $s = 1$, $r = k$, $t = 1$) yields a bound of $2a(\gamma)\ (k+3)/(n-m)$ for the first term in (3.2). Proposition A.4 gives the upper bound

$$2[(\det\pi Q\pi')^{-1/2} - 1]$$

for the second term of (3.2). The completes the proof. □

<u>Example 3.1 continued</u>: Here we continue the discussion of simple linear regression. The notation established in Example 3.1 is used here. Theorem 3.1 shows that the variation distance between the distribution of Y_1 and the closest mixture of normals is bounded above by

$$B_{k,n}(X) = 2a(\gamma)\frac{k+3}{n-2} + 4\left[\left(\det\pi' Q\pi\right)^{-1/2} - 1\right],$$

as long as $(k + 3) \leq \gamma(n-2)$. For this example,

$$\pi = (I_k\ 0)\ :\ k{\times}n$$

and

$$Q = I_n - X(X'X)^{-1}X'$$

since X has full rank. Therefore

$$\pi' Q \pi = I_k - \pi X(X'X)^{-1}X'\pi' =$$

$$I_k - X_1(X_1'X_1 + X_2'X_2)^{-1}X_1'$$

where

$$X = \begin{pmatrix} X_1 \\ \\ X_2 \end{pmatrix} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}$$

A bit of algebra shows that

$$\det \pi' Q \pi = \frac{\det(X_2'X_2)}{\det(X_1'X_1 + X_2'X_2)}$$

It is shown in Lauritzen (1982, p. 225-6) that

$$(3.3) \qquad \lim_{n \to \infty} [(\det \pi' Q \pi)^{-1/2} - 1] = 0$$

iff

$$(3.4) \qquad \lim_{n \to \infty} \sum_1^n (t_i - \bar{t})^2 = +\infty \ .$$

Thus, for k and γ fixed, $B_{k,n}(X)$ converges to 0 as $n\text{-->}\infty$ iff (3.4) holds.

Here is a statistical interpretation of condition (3.3). Because all the eigenvalues of $\pi'Q\pi$ are less than or equal to one, it is clear that (3.3) holds iff

$$\text{(3.5)} \qquad \lim_{n\text{-->}\infty} \pi'Q\pi = \lim_{n\text{-->}\infty} (I_k - X_1(X'X)^{-1}X_1') = I_k$$

But (3.5) holds iff

$$\text{(3.6)} \qquad \lim_{n\text{-->}\infty} (X'X)^{-1} = 0,$$

because X_1 is fixed in our discussion. However, in the regression problem with n observations, the least squares estimator of β, say $\hat{\beta}_n$, satisfies

$$\mathrm{Cov}(\hat{\beta}_n) = \sigma^2(X'X)^{-1}$$

where Cov(•) denotes covariance matrix. Here σ^2 is the common variance of the errors. Thus (3.6) holds iff $\hat{\beta}_n$ is a consistent estimator of β. □

Appendix

Several technical results are collected in this appendix. These include some distributional results about random orthogonal matrices and variation distance bounds.

Let M be a subspace of R^n and define the compact group O(M) by

$$O(M) = \{g \mid g \in O_n,\ gx = x \text{ for } x \in M\} \tag{A.1}$$

As usual, O_n denotes the group of n×n orthogonal matrices. Let ν denote the unique invariant probability measure on O(M). A random matrix $U \in O(M)$ is <u>uniform on O(M)</u> if $\mathcal{L}(U) = \nu$. In order to simplify certain calculations, it is useful to observe that for any fixed $\Gamma \in O_n$,

$$O(\Gamma M) = \Gamma O(M)\Gamma'. \tag{A.2}$$

<u>Lemma A.1</u>: Let $m = \dim M$, $0 \le m < n$ and let P be the orthogonal projection onto M. Consider the random matrix U, uniform on O(M), as an element of the inner product space $\mathcal{L}_{n,n}$ of n×n matrices endowed with the trace inner product. In this inner product space, the mean and covariance of U are

$$\begin{cases} \mathcal{E}U = P \\ \mathrm{Cov}(U) = (n-m)^{-1}Q\otimes Q \end{cases} \tag{A.3}$$

where $Q = I-P$ is the orthogonal projection onto $M^{\perp}$, and $\otimes$ denotes the Kronecker product (as defined in Eaton (1983), p. 34; and also see Chapter 2 of Eaton (1983) for the description of covariances of matrices using the Kronecker product).

Proof: (A.3) is first established for the special subspace

$$(A.4) \qquad M_o = \{x \mid x \in R^n,\ x = \begin{pmatrix} \dot{x} \\ 0 \end{pmatrix},\ \dot{x} \in R^m\}$$

Then (A.2) is used to obtain the general result. For M_o, observe that

$$O(M_o) = \left\{ g \mid g \in O_n,\ g = \begin{pmatrix} I_m & 0 \\ 0 & g^* \end{pmatrix},\ g^* \in O_{n-m} \right\}.$$

Thus, if U^*: $(n-m)\times(n-m)$ is uniform on O_{n-m}, then

$$U_o = \begin{pmatrix} I_m & 0 \\ 0 & U^* \end{pmatrix}$$

is uniform on $O(M_o)$. Because $\mathcal{L}(U^*) = \mathcal{L}(-U^*)$, $U^* = 0$ so

$$\mathcal{E}U_o = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} = P_o$$

where P_o is the orthogonal projection onto M_o. Thus the first relation in (A.3) holds when $M = M_o$. To verify the second relation, it suffices to show that

$$(A.5) \qquad \mathrm{var}(\mathrm{tr}\ A\ U_o) = \mathrm{tr}\ A\ (\ (n-m)^{-1}\ Q_o \otimes Q_o)\ A'$$

for each nxn real matrix A (see Proposition 2.6 in Eaton (1983), where

$$Q_o = I - P_o = \begin{pmatrix} 0 & 0 \\ 0 & I_{n-m} \end{pmatrix}$$

Partition A as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

with A_{11}: mxm and A_{22}:(n-m)x(n-m).

Then

$$\text{var}(\text{tr } A\ U_o) = \text{var}\ (\text{tr } A_{11} + \text{tr } A_{22}\ U^*) =$$
$$\text{var}(\text{tr } A_n\ U^*).$$

Because the distribution of U* is invariant under right and left orthogonal transformations, two applications of Proposition 2.19 in Eaton (1983) show that

$$\text{Cov}\ (U^*) = c\ I_{n-m} \otimes I_{n-m}$$

where the constant c is the variance of any element of U*. That $c = (n-m)^{-1}$ follows from a routine calculation. Hence

$$\text{var}(\text{tr } A\, U_o) = \text{var}(\text{tr } A_{22}U^*) =$$

$$\text{tr } A_{22}[(n-m)^{-1}I_{m-m} \otimes I_{n-m}]A'_{22} =$$

$$(n-m)^{-1}\text{tr } A_{22}A'_{22}.$$

But, the rhs of (A.5) is

$$(n-m)^{-1}\text{tr } A\, (Q_o \otimes Q_o)A' =$$

$$(n-m)^{-1}\text{tr } A\, Q_o A' Q_o =$$

$$(n-m)^{-1}\text{tr } A_{22}A'_{22}.$$

Thus, (A.3) holds when $M = M_o$. For an arbitrary subspace M of dimension m, there exists a tranformation $\Gamma \epsilon O_n$ such that $\Gamma M = M_o$. Then (A.2) yields

$$O(M_o) = O(\Gamma M) = \Gamma O(M)\, \Gamma'$$

Thus, if U_o is uniform on $O(M_o)$, $\Gamma' U_o \Gamma = U$ is uniform on $O(M)$.

Therefore

$$\mathcal{E}U = \Gamma' U_o \Gamma = \Gamma'(\mathcal{E}U_o)\, \Gamma =$$

$$\Gamma' P_o \Gamma = P$$

where P is the orthogonal projection onto M. Similarly,

$$\mathrm{Cov}\,(U) = \mathrm{Cov}(\Gamma' U_o \Gamma) =$$
$$\mathrm{Cov}[(\Gamma'\otimes\Gamma')U_o] = (\Gamma'\otimes\Gamma')\ \mathrm{Cov}(U_o)(\Gamma\otimes\Gamma) =$$
$$(n-m)^{-1}(\Gamma'\otimes\Gamma')\ (Q_o\otimes Q_o)(\Gamma\otimes\Gamma) =$$
$$(n-m)^{-1}(\Gamma' Q_o \Gamma)\otimes(\Gamma' Q_o \Gamma) = (n-m)^{-1}\ Q\otimes Q$$

where $Q = I - P$. □

Again let U be uniform on $O(M)$, let P be the orthogonal projection onto M and set $Q = I - P$. For two given matrices α: $r\times n$ and β: $s\times n$, the results below describe the distribution of

(A.6) $$V = (\alpha\otimes\beta)U = \alpha U \beta' \quad .$$

With $m = \dim M < n$, it is assumed that r and s are no larger than $n - m$. The two non-negative definite matrices

(A.7) $$A = \alpha Q \alpha' \quad , \ B = \beta Q \beta'$$

appear below as do their symmetric squre roots denoted by $A^{1/2}$ and $B^{1/2}$.

<u>Proposition A.1</u>: Let U^* be uniform on O_{n-m} and let Δ: $r\times s$ be the upper left corner block of U^*. Then

(A.8) $$\mathcal{L}(V) = \mathcal{L}(A^{1/2}\Delta B^{1/2} + \alpha P\beta') .$$

Further, if $r+s \le n-m$, then Δ has a density (with repsect to Lebesgue measure) concentrated on the set $\{\Delta'\Delta \le I_s\}$. When $s \le r$, the density of Δ is

(A.9) $$f(\Delta;r,s) = (\sqrt{2\pi})^{-rs} \frac{w(n-m-r,s)}{w(n-m,s)} |I_s-\Delta'\Delta|^{(n-m-r-s-1)/2}$$

where $w(\cdot,\cdot)$ is the Wishart constant defined by

(A.10) $$[w(t,p)]^{-1} = \pi^{p(p-1)/4} \quad 2^{tp/2} \prod_{j=1}^{p} \Gamma\left(\frac{t-j+1}{2}\right) .$$

Here p is a positive integer and t is a real number, $t > p-1$. When $r \le s$, the density of Δ is obtained by interchanging r and s in the Wishart constants in (A.9).

<u>Proof</u>: As in Lemma A.1, it suffices to establish the proposition when $M = M_o$ given in (A.4). In this case

$$P_o = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}$$

is the projection onto M_o and

(A.11) $$Q_o = (I - P_o) = C_oC_o'$$

where

(A.12) $$C_o = \begin{pmatrix} 0 \\ I_{n-m} \end{pmatrix} : n\times(n-m) .$$

When $M = M_o$

$$V = \alpha\, U_o\beta' = \alpha(P_o+Q_o)U_o(P_o+Q_o)\, \beta' =$$

$$\alpha\, Q_oU_oQ_o\beta' + \alpha P_o\beta' .$$

The third equality follows from $P_oQ_o = 0$ and $P_oU_o = U_oP_o = P_o$.

As in the proof of Lemma A.1,

$$U_o = \begin{pmatrix} I_m & 0 \\ 0 & U^* \end{pmatrix}$$

where U^* is uniform on O_{n-m}. The use of (A.11) now yields

(A.13) $$V = \alpha C_oC_o'U_oC_oC_o'\beta' + \alpha P_o\beta' = \gamma\, U^*\delta' + \alpha P_o\beta'$$

where

(A.14) $$\gamma = \alpha\, C_o \text{ and } \delta = \beta C_o \quad .$$

With

$$A_o = \gamma\, \gamma' = \alpha Q_o \alpha',\ B_o = \delta\delta' = \beta Q_o \beta',$$

γ and δ can be written

(A.15) $$\gamma = A_o^{1/2}(I_r 0)\ \psi_1,\ \delta = B_o^{1/2}(I_s 0)\psi_2$$

where $\psi_i \epsilon\ O_{n-m}$, i=1,2. This well-known representation (sometimes call the polar decomposition) follows easily from Vinograd's Theorem (see Eaton, (1983), Example 1.11, p. 37). Because U^* is uniform on O_{n-m},

$$\mathcal{L}(U^*) = \mathcal{L}(\psi_1\ U^* \psi_2').$$

Substituting (A.15) into (A.13) yields

$$\mathcal{L}(V) = \mathcal{L}(A_o^{1/2}(I_r 0)\psi_1 U^* \psi_2' \begin{pmatrix} I_s \\ 0 \end{pmatrix} B_o^{1/2} + \alpha P_o \beta') =$$

$$\mathcal{L}(A_o^{1/2}(I_r 0)U^* \begin{pmatrix} I_s \\ 0 \end{pmatrix} B_o^{1/2} + \alpha P_o \beta') =$$

$$\mathcal{L}(A_o^{1/2}\ \Delta\ B_o^{1/2} + \alpha\ P_o \beta')$$

where

$$\Delta = (I_r 0)\ U^* \begin{pmatrix} I_s \\ 0 \end{pmatrix}$$

is the r×s upper left block of U^*. Thus (A.8) holds when $M = M_o$. For general M, just repeat the argument given in the proof of Lemma A.1.

The second assertion concerning the density of Δ, originally due to Khatri (1970), is proved via invariance methods in Eaton (1985). □

Corollary A.1: If $r+s \le n-m$ and if A and B have full rank, then V has a density given by

$$f(V) = |A|^{-s/2}|B|^{-r/2}\ f(A^{-1/2}(V-\alpha P\beta')B^{-1/2};r,s) \tag{A.16}$$

where $f(\cdot;\ r,s)$ is given by (A.9).

Proof: Using (A.8) and computing a Jacobian, the result follows immediately. □

Now, suppose Δ:r×s has the density (A.9) with $m = 0$, and $r+s \le n$. Thus Δ is the r×s left upper block of a random matrix U on O_n. Clearly $E\Delta=0$ and and

$$\text{Cov}(\Delta) = 1/n\ I_r \otimes I_s.$$

Also let X:r×s have a multivariate normal distribution with the same mean and covariance as Δ. With $\mathcal{L}(X) = P_1$ and $\mathcal{L}(\Delta) = P_2$, the results below give an upper bound on the variation distance

$$\delta_{r,n} = ||P_1 - P_2|| = 2 \sup_B |P_1(B) - P_2(B)|$$

between P_1 and P_2. Here B ranges over all Borel sets. In what follows, the case of $r \geq s$ is treated. In this case the density of X is

$$\text{(A.17)} \qquad f_1(x) = (\sqrt{2\pi})^{-rs} n^{rs/2} \exp\,[-1/2\; n \;\mathrm{tr}\; x'x]$$

where x is an $r \times s$ real matrix.

The density of Δ is

$$\text{(A.18)} \qquad f_2(x) = (\sqrt{2\pi})^{-rs} \frac{w(n-r,s)}{w(n,s)} |I_s - x'x|^{(n-r-s-1)/2} I_o(x'x)$$

where I_o is given by

$$I_o(x'x) = \begin{cases} 1 & \text{if } 0 < x'x \leq I_s \\ 0 & \text{otherwise} \end{cases} \quad .$$

Because f_1 and f_2 are both functions of $x'x$, the variation distance $\delta_{r,n}$ is equal to the variation distance between the distribution of $x'x$ and of $\Delta'\Delta$. The density function of $x'x$ is

$$\text{(A.19)} \qquad g(v) = w(r,s) n^{rs/2} |v|^{(r-s-1)/2} \exp[-1/2n\; \mathrm{tr}\; v]\; J_o(v)$$

where v is a s×s symmetric matrix and

$$J_o(v) = \begin{cases} 1 & \text{if } v > 0 \\ 0 & \text{otherwise} \end{cases} \quad .$$

The density function of $\Delta'\Delta$ is

(A.20) $$g_2(v) = \frac{w(r,s)\ w(n-r,s)}{w(n,s)} |v|^{(r-s-1)/2} |I-v|^{(n-r-s-1)/2} I_o(v).$$

This multivariate beta density can be found a number of places for example, see Olkin and Rubin (1964). The variation distance $\delta_{r,n}$ is thus

$$\delta_{r,n} = \int (g_2(v) - g_1(v)|dv = 2 \int_E \left(\frac{g_2(v)}{g_1(v)} - 1 \right) g_1(v)dv$$

where E is the set p×p positive definite matrices such that $g_2(v) > g_1(v)$. Hence

(A.21) $$1/2\ \delta_{r,n} \leq \sup_{v \in E} \left[\frac{g_2(v)}{g_1(v)} - 1 \right] = M_{r,n} \quad .$$

Differentiation shows that the sup in (A.21) is achieved uniquely for v equal to $\hat{v} = (r+s+1)n^{-1}I_s$. After some algebra this yields

(A.22) $M_{r,n} + 1 =$

$$\prod_{j=1}^{s}\left[\frac{\Gamma\left(\frac{n-j+1}{2}\right)}{\Gamma\left(\frac{n-r-j+1}{2}\right)\left(\frac{n}{2}\right)^{r/2}}\exp\left\{\frac{n-r-s-1}{2}\log\left(1-\frac{r+s+1}{n}\right)+\frac{r+s+1}{2}\right\}\right]$$

We now proceed to bound $M_{r,n}+1$ when r is an even integer. The case of r odd is treated later. For notational covenience, let

$$\text{(A.23)}\qquad \begin{cases} m = n/2 \\ p = \frac{r}{2}+1 \\ t = (r+s+1)/2 \\ m_j = (n-j+1)/2 \end{cases} .$$

Then

$$\text{(A.24)}\qquad M_{r,n}+1 = \prod_{j=1}^{s}\left[\exp\left\{\log\frac{\Gamma(m_j)}{\Gamma(m_j-p+1)m^{p-1}} + m\int_0^{t/m}-\log(1-x)\,dx\right\}\right]$$

Because r is even, p is an integer so

$$\text{(A.25)}\qquad \log\frac{\Gamma(m_j)}{\Gamma(m_j-p+1)m^{p-1}} = \log\prod_{i=1}^{p-1}\left(\frac{m_j-i}{m}\right) =$$

$$\sum_{i=1}^{p-1}\log\left(1-\frac{i+(j-1)/2}{m}\right)$$

Now, observe that $x \to -\log(1-x)$ is an increasing convex function on [0,1] so that for $0 \le a \le b < 1$,

(A.26) $$\int_a^b -\log(1-x)dx \le \frac{(b-a)}{2} \left[-\log(1-b) - \log(1-a)\right] .$$

Using (A.26) yields the inequality

(A.27) $$\log\left(1 - \frac{i+(j-1)/2}{m}\right) \le$$

$$\frac{1}{2}\left[\log\left(1 - \frac{i+(j-1)/2}{m}\right) - \log\left(1 - \frac{i-1+(j-1)/2}{m}\right)\right]$$

$$+ m \int_{\frac{i-1+(j-1)/2}{m}}^{\frac{i+(j-1)/2}{m}} \log(1-x)dx$$

Thus,

(A.28) $$A_j = \sum_{i=1}^{p-1} \log\left(1 - \frac{i+(j-1)/2}{m}\right) + m \int_0^{t/m} -\log(1-x)dx \le$$

$$\frac{1}{2}\left[\log\left(1 - \frac{p-1+(j-1)/2}{m}\right) - \log\left(1 - \frac{(j-1)/2}{m}\right)\right] +$$

$$m \int_0^{\frac{(j-1)/2}{m}} -\log(1-x)dx + m \int_{\frac{p-1+(j-1)/2}{m}}^{t/m} -\log(1-x)dx \quad .$$

Using (A.26) on the final two integrals in (A.28) leads to the inequality

$$\text{(A.29)} \qquad A_j \le \frac{1}{2}\left(\frac{j+1}{2}\right)\left(-\log\left(1-\frac{j-1}{2m}\right)\right) +$$

$$\frac{1}{2}\left(t - 1 - \left(p-1+\frac{j-1}{2}\right)\right)\left[-\log\left(1-\frac{p-1+(j-1)/2}{m}\right)\right] +$$

$$\frac{1}{2}\left(t - \left(p-1+\left((j-1)/2\right)\right)\right)\left[-\log\left(1-t/m\right)\right] \; .$$

Since the index j ranges from 1 to s,

$$\text{(A.30)} \qquad \begin{cases} -\log\left(1-\frac{j-1}{2m}\right) \le -\log\left(1-\frac{s-1}{2m}\right) \\ -\log\left(1-\frac{p-1+(j-1)/2}{m}\right) \le -\log(1-t/m). \end{cases}$$

Using (A.30) on (A.29) and summing on j yields

$$\text{(A.31)} \qquad \sum_{j=1}^{s} A_j \le \frac{s(s+3)}{8}\left[-\log\left(1-\frac{s-1}{2m}\right)\right] +$$

$$\frac{s(s+1)}{4}\left[-\log\left(1-t/m\right)\right] \; .$$

Since $s-1 \le t$, (A.31) gives

$$\text{(A.32)} \qquad \sum_{j=1}^{s} A_j \le \frac{3s^2 + 5s}{8} \left[-\log\left(1 - \frac{t}{m}\right)\right] =$$

$$\frac{3s^2 + 5s}{8} \left[-\log\left((1 - \frac{r+s+1}{n}\right)\right] .$$

This inequality yields

<u>Proposition A.2</u>: Assume $\Delta{:}r\times s$ has the density (A.9) with $m=0$ and $r+s \le n-2$. Let $X{:}r\times s$ be multivariate normal with the same mean and covariance as Δ. When $s \le r$ and r is even, the variation distance between $\mathcal{L}(\Delta)$ and $\mathcal{L}(X)$, say $\delta_{r,n}$, satisfies

$$\text{(A.33)} \qquad 1/2\delta_{r,n} \le \exp\left\{\left(\frac{3s^2+5s}{8}\right)\left[-\log\left(1 - \frac{r+s+1}{n}\right)\right]\right\} -1$$

<u>Proof</u>: Since

$$1/2\delta_{r,n} + 1 \le M_{r,n} + 1 \le \prod_{j=1}^{s} \exp(A_j),$$

inequality (A.32) gives (A.33). □

To treat the case when r is odd, first observe that

$$\text{(A.34)} \qquad \delta_{r,n} \le \delta_{r+1,n}.$$

Inequality (A.34) is a consequence of the easily verified inequality

$$(A.35) \qquad ||\mathcal{L}(h(U_1)) - \mathcal{L}(h(U_2))|| \le ||\mathcal{L}(U_1) - \mathcal{L}(U_2)||$$

which is valid for any U_1 and U_2 and measurable function h. To verify (A.34) just pick $U_1 = \Delta:(r+1)\times s$ and $X:(r+1)\times s$ to be $N(0, n^{-1}I_{r+1}\otimes I_s)$. Then the rhs of (A.35) is $\delta_{r+1,n}$. Pick h to be the linear transformation which discards the last row of an $(r+1)\times s$ matrix. Then the lhs of (A.35) is $\delta_{r,n}$. Thus (A.34) holds and we have

<u>Corollary A.1</u>: For $r+s \le n-3$ and $t = \min(r,s)$, the variation distance between $\mathcal{L}(\Delta)$ and $\mathcal{L}(X)$ satisfies

$$(A.36) \qquad 1/2\delta_{r,n} \le \exp\left\{\frac{3t^2 + 5t}{8}\left[-\log\left(1 - \frac{r+s+2}{n}\right)\right]\right\} - 1 .$$

<u>Proof</u> First assume $s \le r$. For r even (A.33) implies (A.36). For r odd,

$$1/2\delta_{r,n} \le 1/2\ \delta_{r+1,n}.$$

But by (A.33), the rhs of (A.36) is an upper bound for $1/2\delta_{r+1,n}$.

When $r \le s$, just interchange the roles of r and s in Proposition A.2. □

When $s = 1$, (A.36) is the bound given in Theorem 5.1 in Diaconis and Freedman (1986). A somewhat simpler upper bound than (A.36) can be given.

<u>Proposition A.3</u> Let $c = (3t^2 + 5t)/8$ with $t = \min\{r,s\}$ and fix a number γ in

(0,1). For $r+s \le n-3$, $(r+s+2)/n \le \gamma$ and

$$a = a(\gamma) = \frac{(1-\gamma)^{-c} - 1}{\gamma} ,$$

$$1/2\ \delta_{r,n} \le a\ \frac{r+s+2}{n} . \tag{A.37}$$

Proof: With $v = (r+s+2)/n$, (A.36) yields

$$1/2\ \delta_{r,n} \le (1-v)^{-c} - 1.$$

Because $c \ge 1$, the function

$$v \dashrightarrow (1 - v)^{-c} - 1$$

is convex. Since the linear function av and $(1 - v)^{-c} - 1$ agree at 0 and γ, (A.37) follows. □

The following result provides a useful upper bound on the variation distance between certain normal distributions. The setting for this result is in a finite dimensional inner product space because this generality is needed in our applications. Let $(V,(\cdot,\cdot))$ be a finite dimension inner product so that the density of a $N(\mu, \Sigma)$ distribution on $(V, (\cdot,\cdot))$ is

$$h(x) = (\sqrt{2\pi})^{-n/2}|\Sigma|^{-1/2}\exp\left[-\tfrac{1}{2}\ (x-\mu,\ \Sigma^{-1}[x-\mu])\right]$$

where $n = \dim V$ and $|\Sigma|$ denotes the determinant of Σ. The dominating measure is Lebesgue measure on V, and of course Σ is assumed to be non-singular.

<u>Proposition A.4</u>: Suppose X is $N(0,I)$ and Y is $N(0,\Sigma)$ on $(V,(\cdot,\cdot))$. If $\Sigma \leq I$, then the variation distance between $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ is bounded above by

$$2(|\Sigma|^{-1/2}-1). \tag{A.38}$$

<u>Proof</u>: Let h_1 be the density for $\mathcal{L}(X)$ and h_2 be the density for $\mathcal{L}(Y)$. Then

$$||\mathcal{L}(X) - \mathcal{L}(Y)|| = \sup_{|f|\leq 1}\left|\int f(x)h_1(x)dx - \int f(x)h_2(x)dx\right| \tag{A.39}$$

where the sup ranges over all measurable functions bounded in absolute value by one. But, it is well known that

$$||\mathcal{L}(X) - \mathcal{L}(Y)|| = 2\int_E \left[\frac{h_2(x)}{h_1(x)} - 1\right] h_1(x)dx \tag{A.40}$$

where

$$E = \{x | h_2(x) > h_1(x)\} .$$

Hence

$$||\mathcal{L}(X) - \mathcal{L}(Y)|| \leq 2 \sup_x \left[\frac{h_2(x)}{h_1(x)} - 1\right] .$$

But, for each x,

$$\frac{h_2(x)}{h_1(x)} \leq |\Sigma|^{-1/2}$$

because $\Sigma \leq I$. This completes the proof. $\square$

Bibliography

Dawid, A.P. (1977). Spherical matrix distributions and a multivariate model. J. Roy. Statist. Soc. B, 39, 254-261.

Dawid, A.P. (1978). Extendability of spherical matrix distributions. J. Mult. Anal., 8, 567-572.

Diaconis, P. and Freedman, D. (1980). Finite exchangeable sequences. Ann. Prob 8, 745-764.

Diaconis, P. and Freedman, D. (1984). Partial exchangeability and sufficiency. In Statistics, Applications and New Directions. J.K. Ghosh and J. Ray (Eds.) pp 205-236. Indian Statistical Institute, Calcutta.

Diaconis, P. and Freedman, D. (1987). A doezen deFinetti-style results in search of a theory. Ann. Inst. Henri Poincare, Sup. 87/02 Vol. 23, 397-423.

Eaton, M.L. (1983). Multivariate Statistics: A Vector Space Approach. Wiley, New York.

Eaton, M.L. (1985). On some multivariate distributions and random orthogonal matrices. In Proceedings of Berkeley in Honor of Jerzy Neyman and Jack Kiefer, Vol. II. (Eds.) Lucien M. LeCam and Richard A. Olshen, p. 681-691, Wadsworth, Belmont, CA.

Khatri, G.C. (1970). A note on Mitra's paper "A density free approach to the matrix variate beta distribution." Sankhȳa, Series A, 32, 311-318.

Lauritzen, S. (1982). Statistical Models as Extremal Families. Aalborg University Press, Aalborg, Denmark.

Olkin, I. and Rubin, H. (1964). Multivariate beta distribution and independence properties of the Wishart distribution. Ann. Math. Statist. 35, 261-269.

Smith, A.F.M. (1981). On random sequences with centered spherical symmetry. J. Roy. Statist. Soc. B., 208-209.